Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.
Abstract

ATM networks are intended to provide a "one-size-fits-all" solution to a variety of data communication needs, from low speed, delay-insensitive to high-speed, delay-intolerant. The basic ATM protocol certainly delivers traffic within this broad range, but it does not address the quality of service requirements associated with the various type of traffic.

The ATMSWITCH is designed to use two different mechanisms to provide the quality of service for the various type of traffic. It treats the cells according to their connected virtual channel type and services them as predefined scheme.

The ATMSWITCH architecture is a shared-memory and output buffer strategy switch. The switch has been designed much of buffer location and identification can occur in parallel with the 12ns read/write cycle time required to buffer the cell data. The problem is essentially one of design circuitry so that buffer location and identification are as short as possible.

The present project has therefore been intended to measure the number of clock cycles required to perform the buffer maintenance activities, and to determine whether the logic speed required to fit this number of clock cycles into the 12ns window is feasible using current technology. The simulated result and timing analysis shows that 10 clock cycles are required during 12ns buffer read and write time, and a reasonable clock speed is 1.2ns per clock cycle.
Acknowledgements

I would like to thank my supervisor Mr. Paul Lyons for his guidance, encouragement and great help during my study.

Thanks also extend to all staff and postgraduate students in the Department of Computer Science for their assistance and friendship.

On the personal side, I would especially like to thank my wife, Wei Ma, and my daughter, Angela for their continued support and encouragement.
# Table of Contents

Abstract
Acknowledgments
Table of Contents

## Chapter 1. Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Circuit Switching</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Packet Switching</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Statistical Multiplexing and The Jitter Problem</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>ISDN and ATM</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>The Problems to be Solved by Cell Relay</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>The Technique of Cell Relay</td>
<td>9</td>
</tr>
<tr>
<td>1.7</td>
<td>ATM Switch</td>
<td>10</td>
</tr>
<tr>
<td>1.8</td>
<td>ATM SWITCH Architecture</td>
<td>12</td>
</tr>
</tbody>
</table>

## Chapter 2. ATM Protocol Stack

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>ATM Concept</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>The Protocol Reference Mode</td>
<td>16</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The physical Layer</td>
<td>18</td>
</tr>
<tr>
<td>2.2.2</td>
<td>The ATM Layer and ATM cells</td>
<td>18</td>
</tr>
<tr>
<td>2.2.3</td>
<td>The ATM Cell Header</td>
<td>18</td>
</tr>
<tr>
<td>2.2.4</td>
<td>General ATM operation</td>
<td>20</td>
</tr>
<tr>
<td>2.2.5</td>
<td>The Four Classes of ATM Traffic</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>ATM Networking</td>
<td>23</td>
</tr>
<tr>
<td>2.3.1</td>
<td>ATM-Layer Cell Transport</td>
<td>23</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Virtual Channels and Virtual Paths</td>
<td>23</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Multiplexors, Switches and Cross-Connects</td>
<td>24</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Controlling The Connections</td>
<td>25</td>
</tr>
<tr>
<td>2.3.5</td>
<td>ATM Switching</td>
<td>26</td>
</tr>
</tbody>
</table>

## Chapter 3. ATM Switch Architecture Overview

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>ATM Switch Definition and Functionality</td>
<td>27</td>
</tr>
</tbody>
</table>

II
Chapter 3. ATM Switch Architecture and Implementation

3.2. Performance Measure
3.3. Traffic Model
3.4. Buffering Strategies
  3.4.1 Output Buffering
  3.4.2 Input Buffering
3.5. ATM Switch Architecture
  3.5.1 Shared-Memory ATM Switches
  3.5.2 Shared-Medium Switches
  3.5.3 Space-Division Switches
3.6. Trigger Mechanism and Chandelier Structure based ATM Switch Architecture
  3.6.1 Architecture Description
  3.6.2 Associative Chandelier Structure
  3.6.3 Trigger Mechanism
  3.6.4 Speed Achievable of Trigger Mechanism and Chandelier Based Switch

Chapter 4. ATM Switch Structure and Implementation

4.1. Introduction ATMSWITCH Structure and its Circuit Modules
4.2. Operation of ATMSWITCH
4.3. Clock Module
4.4. Input Port Module
4.5. Input Round Robin Server Module
4.6. Counter Module
4.7. Controller Module
4.8. Map Memory
  4.8.1 Structure of a Word in Map Memory
  4.8.2 Chandelier Structure
  4.8.3 Implementation of Map Memory
4.9. Trigger Mechanism
  4.9.1 Principle of Trigger Mechanism
  4.9.2 ITV Space and Mask
  4.9.3 Implementation of Trigger Mechanism
4.10. Cell Memory Module
4.11. Cell Queue Link Module
4.12. Output Round Robin Server
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.13. Output Port</td>
<td>88</td>
</tr>
<tr>
<td>4.14. Control Algorithm</td>
<td>89</td>
</tr>
<tr>
<td>4.15. Summary</td>
<td>94</td>
</tr>
<tr>
<td>Chapter 5. Simulating Result and Performance Analysis</td>
<td>95</td>
</tr>
<tr>
<td>5.1. Simulating Result</td>
<td>96</td>
</tr>
<tr>
<td>5.1.1 Accepting Cell</td>
<td>96</td>
</tr>
<tr>
<td>5.1.2 Outgoing Cell</td>
<td>106</td>
</tr>
<tr>
<td>5.1.3 Clock Speed and Switch Working Cycle</td>
<td>116</td>
</tr>
<tr>
<td>5.2. Cell Memory Size Requirement</td>
<td>117</td>
</tr>
<tr>
<td>5.2.1 Cell Memory Size for Simple Traffic Situation</td>
<td>117</td>
</tr>
<tr>
<td>5.2.2 Cell Memory Size for Output Overload</td>
<td>119</td>
</tr>
<tr>
<td>5.3. Throughput</td>
<td>121</td>
</tr>
<tr>
<td>5.4. Map Memory Size</td>
<td>122</td>
</tr>
<tr>
<td>5.5. Summary</td>
<td>122</td>
</tr>
<tr>
<td>Chapter 6. Future Work</td>
<td>123</td>
</tr>
<tr>
<td>6.1. Multi-Buffer Architecture</td>
<td>123</td>
</tr>
<tr>
<td>6.2. General Operation of Multi-Buffer Architecture</td>
<td>125</td>
</tr>
<tr>
<td>6.3. Summary</td>
<td>126</td>
</tr>
<tr>
<td>Chapter 7. Conclusions</td>
<td>127</td>
</tr>
<tr>
<td>References</td>
<td>129</td>
</tr>
<tr>
<td>Appendix VHDL Program Listing</td>
<td>132</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

In the past, separate telecommunications networks have been specially designed for different services. For instance, the public switched telephone network has been developed for conversational speech; data networks for computer communications; and broadcast networks for television. While these networks are very capable of supporting their intended services, they are generally not well suited for other services that are inherently different in such network requirements as bandwidth, holding times, end-to-end delays, and error rates.

The idea of a single ubiquitous network providing all services in an integrated manner has existed for some time. In a sense, this notion was suggested by AT&T's first president, Theodore Vail, in his vision of "one policy, one system, universal service" [Boettinger, 1983], and it has been recently restated [Mayo, 1985]. Implementation has been hindered by the lack of the necessary technology and public demand from multimedia applications, which involve the processing and exchange of information in the various media of text, audio, and images. In addition, research in high-speed switching, fibre optics, and new protocols have demonstrated the technical feasibility of integrated services networks.

Research in the past few years has led to the concept of ATM as a possible means to realising an integrated services network. The ATM will represent a dramatic change in the evolution of the public switched telephone network. It combines characteristics of both conventional packet switching and circuit switching. Not only does ATM imply a major change in the network facilities, but for the first time, the network will be designed to provide much more than POTS (Plain Old Telephone Services).

In order to explain how the ATM approach combines both circuit and packet switching techniques to provide all services in an integrated manner, I will first introduce public telephone network and packet-switched data networks. The second part of this chapter will introduces ISDN (Integrated Services Digital Network) and ATM based public switched network that derived from ISDN and promote to B-ISDN. The later part of this chapter will introduce ATMSWITCH that was simulated in this research project.
1.1. Circuit Switching

In traditional analogue circuit switching, a call is set up on the basis that it receives a path (from source to destination) that is its "property" for the duration of the call. I.e. the whole of the bandwidth of the circuit is available to the calling parties for the whole of the call. In a digital circuit-switched system, the whole bit-rate of the line is assigned to a call for only a single time slot per frame. This is called time division multiplexing.

![Diagram of time division multiplexing](image)

**Figure 1.1:** An example of time division, position, multiplexing

During the time period of a frame, the transmitting party will generate a fixed number of bits of digital data (for example, eight bits to represent the level of an analogue telephony signal) and these bits will be grouped together in the time slot assigned in every frame for the duration of the call, to that call (Figure 1.1). So the time slot is identified by its position in the frame, hence use of the name "position multiplexing", although this term is not used as much as time division multiplexing.

When a connection is set up, a route is found through the network and that route remains fixed for the duration of the connection. The route will probably traverse a number of switching nodes and require the use of a similar number of transmission links to provide a circuit from source to destination. The time slot position used by a call is likely to be different on each link. The switches which interconnect the transmission links perform the time slot interchange (as well as the space switching) necessary to provide the
“through-connection” (e.g. link M, time slot 2 switches to link N, time slot 6 in Figure 1.2).

![Diagram](http://example.com/diagram.png)

**Figure 1.2** Time slot interchange

In digital circuit switched telephone networks, frames have a repetition rate of 8000 frames per second (and so a duration of 125 µs), and as there are always eight bits (one byte) per time slot, each channel has a bit-rate of 64 Kbit/s. With \( N \) time slots in each frame, the bit-rate of the line is \( N \times 64 \) Kbit/s. In practice, extra time slots or bits are added for control and synchronisation functions. So for example, the widely used 30 channel system has two extra time slots, giving a total of 32 time slots, and thus a bit-rate of \( (30+2) \times 64 = 2048 \) Kbit/s.

The time division multiplexing concept can be applied recursively by considering a 24 or 30 channel system as a single “channel”, each frame of which occupies one time slot per frame of a higher order multiplexing system. This is the underlying principle in the SDH (Synchronous Digital Hierarchy), and an introduction to SDH can be found in [Griffiths, 1992].

The main performance issue for the user of a circuit switched network is whether, when a call is requested, there is a circuit available to the required destination. Once a circuit is established, the user has available a constant bit-rate with a fixed end-to-end delay. There is no error detection or correction provided by the network on the circuit – that is the responsibility of the terminals at either end, if it is required. Nor is there any per circuit overhead -- the whole bit-rate of the circuit is available for user information.

**1.2. Packet switching**
To see how ATM has evolved from both circuit switched and packet switched networks, it is helpful to consider a "generic" packet switching network. I.e. one intended to represent the main characteristics of packet switching, rather than any particular packet switching system.

Instead of being organized into eight-bit time slots which repeat at regular intervals, data in a packet switched network is organized into packets comprising many bytes of user data (bytes are also known as octets, in order to divorce them from any association with an eight-bit coding scheme). Packets can vary in size depending on how much data there is to send, usually up to some predetermined limit (for example, 4096 octets). Each packet is then sent from switching node to switching node as group of contiguous bits fully occupying the link bit-rate for the duration of the packet. If there is no packet to send, then nothing is sent on the link. When a packet is ready, and the link is idle, then the packet can be sent immediately. If the link is busy (another packet is currently being transmitted), then the packet must wait in a buffer until the previous one has completed transmission (Figure 1.3).

![Figure 1.3: An example of Packet Switching](image)

Each packet has a label to identify it as belonging to a particular communication (The word “communication” has the same notion as the word “call” in the circuit switching. The “communication” is usually used in the data communication term.). Thus packets from different sources and to different destinations can be multiplexed over the same link by being transmitted one after the other. This is also called label multiplexing [Pitts,
The label is used at each node to select an outgoing link, routing the packet across the network. The outgoing link selected may be predetermined at the set-up of the connection, or it may be varied according to traffic conditions (e.g. take the least busy route). The former method ensures that packets arrive in the order in which they were sent, whereas the latter method requires the destination to be able to resequence out-of-order packets (in the event that the delays on alternative routes are different).

Whichever routing method is used, the packets destined for a particular link must be queued in the node prior to transmission. It is this queuing which introduces variable delay to the packets. A system of acknowledgment ensures that corrupted packets are not lost but are retransmitted. This is done on a link-by-link basis, rather than end-to-end, and contributes further to the variation in delay. There is quite a significant per packet overhead required for the error control and acknowledgment mechanisms, in addition to the label. This overhead reduces the effective bit-rate available for the transfer of user information. The packet plus link overhead is often (confusingly) called a “frame”. Note that it is not the same as a frame in circuit switching.

A simple packet-switched network may continue to accept packets without assessing whether it can cope with the extra traffic or not. Thus it appears to be non-blocking, in contrast to a circuit switched network which rejects (blocks) a connection request if there is no circuit available. The effect of this non-blocking operation is that packets experience greater and greater delay across the network, as the load on the network increases. As the load approaches the network capacity, the node buffers become full, and further incoming packets cannot be stored. This triggers retransmission of those packets which only worsens the situation by increasing the load; the successful throughput of packets decreases significantly.

In order to maintain throughput, congestion control techniques, particularly flow control, are used. Their aim is to limit the rate at which sources offer packets to the network. The flow control can be exercised on a link-by-link, or end-to-end basis. Thus a connection cannot be guaranteed any particular bit-rate: it is allowed to send packets to the network as and when it needs to, but if the network is congested then the network exerts control by restricting this rate of flow.

The main performance issues for a user of a packet switched network are the delay experienced on any connection and the throughput. The network operator aims to maximize throughput and limit the delay, even in the presence of congestion. Once a connection is established, the user is able to send information on demand. The network
provides error control through re-transmission of packets on a link-by-link basis. Capacity is not dedicated to the connection, but shared on a dynamic basis with other connections. The capacity available to the user is reduced by the per packet overheads required for label multiplexing, flow and error control.

1.3. Statistical Multiplexing and The Jitter Problem

Packet switching is also called statistical multiplexing. ATM is actually based upon a statistical multiplexing technique called cell relay switching [Clark, 1996]. Statistical multiplexing is a means of multiplying the effective capacity of a transmission line or network, by taking advantage of variations in traffic intensity associated with different calls.

For example, when speech connections are statistically multiplexed, the silent periods can be suppressed and not sent over the line. Meanwhile the words from other conversations can be carried in the gaps.

The major benefit of statistical multiplexing is that the useful carrying capacity of the line is maximized by avoiding the unnecessary transmission of redundant information (i.e. pauses). The first practical realizations of statistical multiplexing were data networking protocols. In particular, statistical multiplexing forms the basis of data packet switching. It is the principle upon which IBM’s SNA (systems network architecture) and ITU-T’s X.25 recommendation are based. As such, statistical multiplexing is widely in use within computer data networks.

Today’s public voice networks, in contrast to data networks, have not used statistical multiplexing. Instead, voice and telephone networks have historically been based upon circuit switching, the allocation of a path across the network on a fully dedicated point-to-point basis for the duration of the call. The strength of circuit switching is the guaranteed throughput and delay performance of the resulting connection. This is critical in order that acceptable voice quality can be achieved (in the subjective opinion of telephone users). A telephone call connected in a circuit-switched manner is like an empty pipe between two telephone users. Whatever one speaker says into the pipe comes out at the other end in an identical format -- but only one pair of callers can use the pipe during any particular call.
Historically, telephone networks have not used statistical multiplexing techniques because of the problem of achieving acceptable speech quality. There were attempts to 'packet switch' voice across data networks, but the problem was that individual words or parts of words take different times to propagate through a packet network, so that the listener hears a rather broken form of the original signal. The effect is caused by a phenomenon called jitter. The more jitter (variable propagation delay) experienced by a telephone connection, the worse the perceived quality of the connection.

Jitter in data networks is relatively unimportant. So long as the average delay is not great, computer users do not notice whether some typed characters appear imperceptibly faster or slower than others. As a result, telephone systems have remained circuit switch based, because of the quality problems. The consequence has been the evolution of two entirely separate networking worlds -- voice and data. Transmission lines cannot easily and efficiently be shared between voice and data, and dynamic allocation of bandwidth -- one instant to voice, the next moment to data -- has not been possible.

1.4. ISDN and ATM

A number of attempts have been made to develop technologies capable of handling equally well both voice and data over the same network. The two most noteworthy technologies in this category are ISDN (integrated services digital network) and, now, ATM (Asynchronous Transfer Mode).

ISDN is a technology based upon circuit switching within digital telephone networks. The digital nature of the network is exploited for the use of data transmission. And to make integrated data and voice carriage possible, the signaling within the network (between the exchanges in the network and from the calling handset to the first exchange) is very advanced -- far superior to the simple pulsing technique used in order analogue telephone networks. Unfortunately, ISDN as a data transport medium is limited in its efficiency and flexibility due to its circuit switched nature.

While ISDN will revamp customer expectations of telephone services (with new features like caller identification before answer and ring back when free), ISDN in its basic form (narrowband ISDN) is unlikely to form the basis of advanced integrated voice and data networks.
In contrast, because ATM (which is a form of so-called broadband ISDN or B-ISDN) has evolved from the statistical multiplexing technique inherent in packet switched data networks, it is likely to have more success as an integrated voice, data and video transport medium. The developers of ATM have simply concentrated on improving the packet switching technique to limit the jitter on speech, video and other delay sensitive applications. The resulting technique is called cell relay switching, or simply cell relay [Clark, 1996].

1.5. The Problems to be Solved by Cell Relay

The normal statistical multiplexing of data connections is carried out by packet switching. Packets of data are created by each of the sources, and interleaved as appropriate by the statistical multiplexor, as illustrated in Figure 1.4:

![Figure 1.4: Statistical multiplexing headers and overhead](image)

The interleaving is usually carried out on a simple FIFO queue basis (first in -- first out). Packets received from the sending sources are stored at the back of the queue. Meanwhile, packets at the front of the queue are being transmitted along the link.

A typical data packet contains between 1 and 256 characters (between 8 and 2048 bits), and the line speed is typically 9600 bit/s. The propagation delay at a time when two sources try to send simultaneously (due to the extra waiting time) may therefore be up to 200 ms (2048/9600 s) longer than when there is no simultaneous transmission. In other words, there may be up to 200 ms of jitter. This is unacceptable for speech transmission. But before discussing how cell relay circumvents this problem, we should cover one
other important aspect of statistical multiplexing -- an aspect which constrains the maximum achievable line usage efficiency.

In order to allow the demultiplexor to sort out the various packets belonging to the different logical connections, and forward them to the correct destinations (A to A, B to B, C to C etc.) there needs to be a label attached to each packet to say to which logical connection (i.e. telephone conversation or data communications session) it belongs. This label is contained in the header, which is an addition to the front of the packet and has a function like the envelope of a letter. The header (Figure 1.4) is crucial to the technique of statistical multiplexing, but has the disadvantage that it adds to the information which must be carried by the transmission line between multiplexor and demultiplexor. At the demultiplexor, the header is removed so it does not disturb the receiver, but meanwhile it has generated an overhead load for the transmission line. It is thus impossible using statistical multiplexing techniques to archive 100 per cent loading of a transmission line with raw user information. Some of the capacity has to be given up to carry the overhead.

The major challenges for ATM developers have therefore been to minimize the jitter experienced by speech, video, and other delay-sensitive applications while simultaneously optimizing line efficiency by minimizing network overhead. As we shall see, these demands contend with one another.

1.6. The Technique of Cell Relay

Cell relay is a form of statistical multiplexing similar in many ways to packet switching, except that the packets are instead called cells. Each of the cells is of a fixed rather than a variable size.

The fixed cell size defined by ATM standards is 48 octets (bytes) plus a 5 octets header (i.e. 53 octets in all -- see Figure 1.5). The transmission line speeds currently foreseen to be used are either 155, 622 or 1200 Mbit/s. We can therefore immediately draw certain conclusions about ATM performance:

- the overhead is at least 5 bytes in 53 bytes, i.e. > 9 per cent;
- the duration of a cell is at most $53 \times 8 \text{ bits/155 Mbit/s} = 2.74 \mu s$. ($0.360 \mu s$ at 1200 Mbit/s).
Since the cell duration is relatively short, provided a priority scheme is applied to allow cells from delay-sensitive signal sources (e.g. speech, video, etc.) to have access to the next cell slot, then the jitter (variation in signal propagation delay) can be kept very low - not zero as is possible with circuit switching, but at least low enough to give a subjectively acceptable quality for telephone listeners or video watchers. Jitter-insensitive traffic sources (e.g. data communication channels) can be made to wait for the allocation of the next free or low priority slot.

Although the cell size and priority scheme can solve the jitter problem in principle, the jitter still could occur when the cell pass the not well-designed switch. So the architecture of an ATM switch is also a key to ensure the quality of time sensitive channel transmission.

1.7. ATM Switch

The cell header carries information sufficient to allow the ATM network to determine to which connection (and thus to which destination port and end-user) each cell should be delivered. We could draw a comparison with a postal service and imagine each of the cells to be a letter with 48 characters of information contained in an envelope on which a 5-digit postcode appears. You simply drop your letters (cells) in the right order and they come out in the same order at the other end, though maybe slightly jittered in time. Just as a postal service has numerous vans, lorries and personnel to carry different letters over different stretches, and sorting offices to direct the letters along their individual paths, so an ATM network can comprise a mesh of transmission links and switches to direct individual cells by inspecting the address contained in the header (Figure 1.6).
The ATM "switch" acts in much the same way as a postal sorter. On its incoming side is a FIFO buffer, like a pile of letters. At the front of the buffer (like the top letter in the pile) is the cell which has been waiting longest to be switched. New cells arriving are added to the back of the buffer. The switching process involves looking at each cell in turn, and determining from the address held in the header which outgoing line should be taken. The cell is then added to the FIFO output buffer which is queuing cells waiting to be transmitted on this line. The cell then proceeds to the next exchange.

ATM will provide flexibility in bandwidth allocation and will allow a network to carry heterogeneous services ranging from narrowband to wideband services requiring real time. However, the challenge is to build fast ATM switches able to match the high speeds of the input links and the high performance requirements imposed.

In 1990, Tobagi described a large number of switching architectures [Tobagi, 1990]. All the approaches point to the need of a very high speed hardware switch because of the high transfer rates involved; on the other hand, due to the statistical multiplexing, buffering is also required in order to avoid cell loss whenever there are multiple input cells arriving simultaneously on different input ports and destined for the same output. Only one cell at a time can be transmitted over an output link; the rest must be temporarily stored in a buffer for later transmission.

An attractive ATM architecture is using shared memory [Hluchyj, 1988] to implement output buffering and therefore to attain the best throughput/delay/cell loss performance.
In shared-memory type switches that operate without blocking, all input and output ports have access to a shared-memory module able to write up to N incoming cells and to read out N outgoing cells in a switching time cycle, so that, as in output-buffered switches, throughput is not reduced by output port contention, and an optimal throughput/delay performance is achieved. But in this architecture, we still need to consider how to design the buffer scheme, how to make jitter low enough when the voice or video cells pass through the switch.

1.8. ATMSWITCH Architecture

ATM networks are intended to provide a "one-size-fits-all" solution to a variety of data communication needs, from low speed, delay-insensitive to high-speed, delay-intolerant. The basic ATM protocol certainly delivers traffic within this broad range, but it does not address the quality of service requirements associated with the various type of traffic. In particular, its statistical multiplexing nature is inherently antagonistic towards constant-delay traffic such as real-time video.

The ATMSWITCH is proposed by Lyons et al in (1996, 1997). The ATMSWITCH uses two different mechanisms to provide the quality of service for the various type of traffic. It treats the cells according to their connected virtual channel type and service them as predefined scheme. The ATMSWITCH ensures that the jitter can be minimized when the CBR/VBR channel’s cell transmission in the ATM network.

The ATMSWITCH architecture is a shared-memory and output buffer strategy switch. On the size the limitations of this type of switch come from the memory control logic (which must be able to handle N incoming and N outgoing cells in each time slot), and the memory bandwidth that must be at least the sum of the bandwidths of the incoming and the outgoing lines. The memory bandwidth depends on the word length, which in turn is limited to the cell size (53 octets = 424b). Therefore, for a given memory cycle time (or memory access time) the number of links N is defined by the following relation (for single ported memories):

\[ N = \frac{cell\_length(b)}{2 \times cycle\_time \times link\_speed(b/s)} \]

Thus a switch incorporating 12ns memory can theoretically support up to 14 1.2 Gbps links. However, this theoretical result does not allow for the other operations which the
switch must perform each time a cell is served - circuit identification, insertion into the buffer, and subsequent extraction from the buffer and assembly into a new cell. Because the buffer structures are comparatively complex in the ATMSWITCH architecture, these operations may consume a significant amount of time. It is therefore important to pack them into the minimum amount of time so that the maximum possible throughput can be achieved. The architecture of the ATMSWITCH has been designed so that much of this administrative business can occur in parallel with the 12ns read/write cycle time required to buffer the cell data. The problem is essentially one of design circuitry so that buffer location and identification and cell assembly are as short as possible and so that the time required for channel identification + buffer insertion is no greater than 12ns, and so that the time required for buffer extraction and cell assembly is no greater than 12ns. This is illustrated in Figure 1.7.

![Figure 1.7: Cache Memory and Switch Timing](image)

The buffer location and identification, and cell assembly operations are minimal complexity - they can be completed in a single clock cycle. Maintenance of the buffers, on the other hand, is quite complex and it was therefore important to minimize the amount of time involved.

The present project has therefore been intended to measure the number of clock cycles required to perform the buffer maintenance activities, and to determine whether the logic speed required to fit this number of clock cycles into the 12ns window is feasible using current technology.

Previous work on the ATMSWITCH has focused on a high-level aspects of the architecture. A detailed logic-level design was not available at the start of this project, and consequently, a large part of the project was the construction of a VHDL (Very High
Speed Integrated Circuit Hardware Description Language) definition of the circuit, so that the number of clock cycles involved could be accurately determined.

An interesting spin-off from the project has been the design of an improved architecture using interleaved memory, which has the potential to improve the speed of the switch still further.